AD-A089 576 SOUTHERN METHODIST UNIV DALLAS TEX DEPT OF STATISTICS F/G 12/1
BIASED ESTIMATION IN REGRESSION.(U)
AUG 80 R F GUNST
UNCLASSIFIED F49620-79-C-0106
AFOSR-TR-80-0865 NL

END
DATE FILMED
10-80
DTIC

(18) AFOSR-TR-80-0865

(19)

(12) P.S.

LEVEL II

(6) BIASED ESTIMATION IN REGRESSION.

AD A089576

DTIC

FINAL REPORT

ON

AIR FORCE OFFICE OF SCIENTIFIC RESEARCH

CONTRACT NO. F49620-79-C-0106 New

(15) (16) 2341

Final rept. 1 Jul 79 - 30 Jun 80,

(17) A5

(10) R... F. Gunst

(12) 12

DEPARTMENT OF STATISTICS

DIVISION OF MATHEMATICAL SCIENCES

SOUTHERN METHODIST UNIVERSITY

DALLAS, TEXAS 75275

DTIC FILE COPY

DTIC ELECTE SEP 30 1980 S D

A

404198

80 9 22 172

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE *(When Data Entered)*

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER AFOSR-TR- 80-0865	2. GOVT ACCESSION NO. AD-A089576	3. RECIPIENT'S CATALOG NUMBER
4. TITLE *(and Subtitle)* BIASED ESTIMATION IN REGRESSION		5. TYPE OF REPORT & PERIOD COVERED Final
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Richard F. Gunst		8. CONTRACT OR GRANT NUMBER(s) F49620-79-C-0106
9. PERFORMING ORGANIZATION NAME AND ADDRESS Southern Methodist University Department of Statistics DAllas, TX 75275		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 61102F 2304/A5
11. CONTROLLING OFFICE NAME AND ADDRESS Air Force Office of Scientific Research/NM Bolling AFB, Washington, DC 20332		12. REPORT DATE August 1980
		13. NUMBER OF PAGES 12
14. MONITORING AGENCY NAME & ADDRESS*(if different from Controlling Office)*		15. SECURITY CLASS. *(of this report)* UNCLASSIFIED
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT *(of this Report)*

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT *(of the abstract entered in Block 20, if different from Report)*

18. SUPPLEMENTARY NOTES

19. KEY WORDS *(Continue on reverse side if necessary and identify by block number)*

20. ABSTRACT *(Continue on reverse side if necessary and identify by block number)*

This report documents the activites of the principal investigator and his collaborators during the contract period 1 July 1979 to 30 June 1980 which were sponsored whole or in part by the AFOSR under Contract F49620-79-C-0106. Summarized in the report are four scholarly papers that have been published during the contract period and four other articles which are in various stages of submission for publication. Several other activities directly connected with this research are briefly discussed. A list of all research articles published since the commencement of the contract concludes the report.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE *(When Data Entered)*

ABSTRACT

This report documents the activities of the principal investigator and his collaborators during the contract period 1 July 1979 to 30 June 1980 which were sponsored whole or in part by the Air Force Office of Scientific Research under Contract No. F49620-79-C-0106. Summarized in the report are four scholarly papers that have been published during the contract period and four other articles which are in various stages of submission for publication. Several other activities directly connected with this research are briefly discussed. A list of all research articles published since the commencement of the present contract and its predecessor, Grant No. AFOSR-75-2871, concludes the report.

Richard F. Gunst
Principal Investigator
Contract No. F49620-79-C-0106
August 1, 1980

TABLE OF CONTENTS

I. RESEARCH ACTIVITIES

A. Summary

Research efforts conducted by the principal investigator and his collaborators during the contract period 1 July 1979 to 30 June 1980 have resulted in the publication of four articles in scholarly journals. Three additional research papers have been completed and are submitted for publication. Finally, one dissertation has been completed under the direction of the principal investigator during the contract period. The major results contained in the dissertation are currently being prepared for submission to a scholarly journal.

Each of the research papers described above is now abstracted.

B. Research Papers Published

1. "An Approach to the Programming of Biased Regression Algorithms," Communications in Statistics, B8, 151-159.

Abstract

Due to the near nonexistence of computer algorithms for calculating estimators and ancillary statistics that are needed for biased regression methodologies, many users of these methodologies are forced to write their own programs. Brute-force coding of such programs can result in a great waste of computer core and computing time, as well as inefficient and inaccurrate computing techniques. This article proposes some guides to more efficient programming by taking advantage of mathematical similarities among several of the more popular biased regression estimators.

2. "Metacercariae of Ornithodiplostomum Ptychocheilus Encysted in the Brains and Viscera of Red-Sided Shiners from the Clark-Fork and Bitterroot Rivers of Montana: An Analysis of the Infected Hosts," Journal of Parasitology, 65, 616-623. (with F. Sogandares-Bernal, and H.J. Hietala)

Abstract

Red-sided shiners (*Richardsonius balteatus*) infected with the metacercariae of *Ornithodiplostomum ptychocheilus* in the brains and viscera were studied to determine if the parasite load had any effect on stamina. A special chamber and electronic circuitry used for these studies are described. Statistical treatment of data derived from the experiments is by multivariate contingency table analysis. Parasites were not found to affect the stamina of the infected fish hosts, and the evolutionary significance of this feature is discussed.

3. "A Relationship Between Generalized and Integrated Mean Squared Errors," Communications in Statistics, A9, 321-326. (with J.L. Hess)

Abstract

Generalized mean squared error is a flexible measure of the adequacy of a regression estimator. It allows specific characteristics of the regression model and its intended use to be incorporated in the measure itself. Similarly, integrated mean squared error enables a researcher to stipulate particular regions of interest and weighting functions in the assessment of a prediction equation. The appeal of both measures is their ability to allow design or model characteristics to directly influence the evaluation of fitted regression models. In this note an equivalence of the two measures is established for correctly specified models.

4. "A Critique of Some Ridge Regression Methods: Comment," Journal of the American Statistical Association, 75, 98-100.

Abstract

Regression analysis, like all statistical procedures, depends on assumptions for its successful implementation. Most data analysts are well aware of the requirements of correct model specification, constant predictor variables, normally distributed errors, and so forth, that underlie classical least squares estimation and inference techniques. Bayesian approaches to a regression analysis incorporate assumptions like these and assumptions about the distribution of the regression coefficients. It is primarily to the debate over assumptions and their reasonableness that Smith and Campbell's paper contributes and that this comment addresses.

C. Additional Completed Research

1. "Evaluation of Prediction Equations When Regression Models are Misspecified," submitted to Technometrics (with J.L. Hess).

Abstract

Integrated mean squared error has been shown to be a valuable criterion for the evaluation of both response surface designs and prediction equations. In this paper the use of integrated mean squared error in the evaluation of alternative prediction equations over a specified region of interest is extended to include misspecified regression models. Least squares, principal component, and ridge regression estimators are investigated for two classes of second-order moment matrices of the weight function. A detailed example provides additional insight into the application of integrated mean squared error in the assessment of prediction equations.

2. "Biased Estimation in Regression: A Ten Year Perspective," submitted to Technometrics.

Abstract

Biased estimation in regression has experienced a tremendous growth in popularity since Hoerl and Kennard's formalization of ridge regression. Along with the interest in biased regression, many research efforts over the last ten years have extended both the theory and application of these methodologies. So too, criticisms have arisen which focus on the incompleteness of the theoretical results and on exaggerated claims about the merits of biased estimators. Rather than attempting to arbitrate these opposing views, this article discusses biased estimation with special emphasis on its ultimate justification: application to real problems. Advantages and disadvantages of three biased estimators (principal component, latent root, and ridge regression estimators) are discussed and illustrated through a comprehensive analysis of a data set on automobile emissions. From the discussion and analysis it is hoped that a more balanced perspective on the application of biased estimation will be fostered.

3. "Residual Analysis: Completing the Picture," submitted to The American Statistician.

Abstract

This article presents three examples of the benefits which can be obtained from a careful, perhaps nonstandard, analysis of residuals. The intent of the paper is to stress the value of studying residuals and the role that curiosity and creativity play in such an analysis. Although it is well-recognized that residual analysis is invaluable in the assessment

of structural questions relating to regression models (e.g., the assumptions of normality, homoscedasticity, independence, etc.), there is a tendency to restrict the examination of residuals to such questions. Recent advances in the application of residual analysis suggest that this viewpoint is far too restrictive. Indeed, residual analysis can supplement and enhance many of the classical estimation and testing procedures as well as help explain apparent anomalies in the data or fit of the model.

4. "Extensions of Inference Procedures for Biased Estimators," in preparation (with T.A. Hua).

Abstract

Two multiparameter biased estimators are discussed in this article: generalized ridge regression and generalized latent root regression. The former family of estimators was defined by Hoerl and Kennard (1970) and developed further by Hocking, Speed, and Lynn (1976); the latter family is defined in the paper. Relationships between the two families are established as are conditions needed for their equivalence. Special cases which reduce the two generalized families of estimators to other well-known biased estimators are also discussed.

II. PRESENTATIONS AND OTHER ACTIVITIES

A. Papers Presented at Professional Meetings

Two formal oral presentations were made by the principal investigator to the scientific community. The first was a contributed paper entitled "Similarities Among Least Squares, Principal Component, and Latent Root Regression Estimators" which was presented at the Joint Annual Meetings of the American Statistical Association and the Biometric Society in Washington, D.C., August 1979.

An invited paper entitled "An Overview of Biased Regression Estimators" was presented at the Seventh Annual I.B.M. Design of Experiments Conference in Austin, Texas, October 1979.

B. Related Professional Activities

1. The principal investigator received the 1979 "Most Outstanding Presentation" Award from the Section on Physical and Engineering Sciences of the American Statistical Association for his presentation at the 1979 Joint Annual Meetings.
2. A doctoral dissertation by T.A. Hua was completed in May 1980 under the direction of the principal investigator.
3. The principal investigator was elected President of the North Texas Chapter of the American Statistical Association.
4. The principal investigator was selected as a recipient of an "Outstanding Young Men of America" award.

III. TITLES OF COMPLETED RESEARCH

This section lists all research papers published since the commencement of Contract No. F49620-79-C-0106 and its predecessor, Grant No. AFOSR-75-2871. Also listed are articles currently being reviewed.

A. Published Papers

1. "Density Functions of the Bivariate Chi-square Distribution," Journal of Statistical Computation and Simulation, 2 (1973), 275-88 (with J.T. Webster).

2. "On Computing Critical Points for a Bivariate Chi-square Random Variable," Communications in Statistics, 2 (1973), 221-24.

3. "Recent Developments in Stepwise Regression Procedures," Proc. of University of Kentucky Conference on Regression with a Large Number of Predictor Variables Oct. 11-12, 1973, Lexington, KY 34-53 (with R.L. Mason and J.T. Webster).

4. "Latent Root Regression Analysis," Technometrics, 16 (1974), 513-22 (with R.L. Mason and J.T. Webster).

5. "The Availability of Tables Useful in Analyzing Linear Models," A Survey of Statistical Design and Linear Models (J.N. Srivastava, ed.), 181-96, North-Holland Publ. Co. (1975) (with D.B. Owen).

6. "Regression Analysis and Problems of Multicollinearity," Communications in Statistics, 4 (1975), 277-92 (with R.L. Mason and J.T. Webster).

7. "A Comparison of Least Squares and Latent Root Regression Estimators," Technometrics, 18 (1976), 75-83 (with R.L. Mason and J.T. Webster).

8. "Reassessing Service Courses in Statistics," Special Issue on Statistical Education, Communications in Statistics, A5 (1976), 925-34, (W.T. Federer, ed.)

9. "Prediction of Mobility Gains in Patients with Cervical Spinal Cord Injuries," Journal of Neurosurgery, 45 (1976), 677-82 (with R.L. Mason).

10. "Generalized Mean Squared Error Properties of Regression Estimators," Communications in Statistics, A5 (1976), 1501-8 (with R.L. Mason).

11. "An Analytic Variable Selection Technique for Principal Component Regression," Applied Statistics, 26 (1977), 34-40 (with E.R. Mansfield and J.T. Webster).

12. "Advantages of Examining Multicollinearities in Regression Analysis," Biometrics, 33 (1977), 249-60 (with R.L. Mason).

13. "Latent Root Regression: Large Sample Analysis," Technometrics, 21 (1979), 481-8 (with J.W. White).

14. "An Approach to the Programming of Biased Regression Algorithms," Communications in Statistics, B8 (1979), 151-9.

15. "Metacercariae of Ornithodoplostomum Ptychocheilus Encysted in the Brains and Viscera of Red-Sided Shiners from the Clark-Fork and Bitterroot Rivers of Montana: An Analysis of the Infected Hosts," Journal of Parasitology, 65 (1974), 616-23 (with F. Sogandares-Bernal and H.J. Hietala).

16. "A Relationship Between Generalized and Integrated Mean Squared Error," Communications in Statistics, A9 (1980), 321-6 (with J.L. Hess).

17. "A Critique of Some Ridge Regression Methods: Comment," Journal of the American Statistical Association, 75 (1980), 98-100.

B. Papers Under Review

1. "Evaluation of Prediction Equations When Regression Models are Misspecified," submitted to Technometrics (with J.L. Hess).

2. "Biased Estimation in Regression: A Ten Year Perspective," submitted to Technometrics.

3. "Residual Analysis: Completing the Picture," submitted to The American Statistician.